Audio Engineering Society

Convention Paper 7701

Presented at the 126th Convention
2009 May 7–10 Munich, Germany

Pitch and Played String Estimation in Classic and Acoustic Guitars

Isabel Barbancho[1], Lorenzo J. Tardón[1], Ana M. Barbancho[1], Simone Sammartino[1]

[1]*E.T.S.I. Telecomunicación, Dpt. Ingeniería de Comunicaciones, Universidad de Málaga, Campus de Teatinos s/n, 29071 Málaga, Spain*

Correspondence should be addressed to Isabel Barbancho (`ibp@ic.uma.es`)

ABSTRACT

In classic and acoustic guitars that use standard tuning, the same pitch can be produced at different strings. The aim of this article is to present a method based on the time and frequency-domain characteristics of the recorded sound to determine, not only the pitch but also the string of the guitar that has been played to produce that pitch. This system will provide information not only of the pitch of the notes played, but also about how those notes were played. This specific information can be valuable to identify the style of the player and can be used in teaching to play the guitar.

1. INTRODUCTION

In both classic and acoustic guitars, there are different possibilities to produce the same pitch depending on the combination of string/fret. Depending on the string/fret combinations selected, a musical piece can be played with different difficulty levels and the sound can be slightly different [1]. In this communication, the time and frequency-domain characteristics of the recorded guitar sound will be analyzed to guide the detection of not only the pitch, but also of the string of the guitar that was plucked. The analysis will be done for both the classic guitar with nylon strings and the acoustic guitar with steel strings, played with different techniques and different dynamics. The outline of the paper is as follows. In section 2, the time and frequency analysis tools employed to extract features in order to estimate the pitch and the the plucked string are presented. The estimation of the plucked string will be called from now on classification. Section 3 presents how the pitch is estimated. Once the pitch has been extracted, the classification problem has to be faced. In section 4, the different features calculated to do the classification are explained and in section 5, the technique selected to take into account a number of features to decide the string plucked is shown. Section 6 presents the results obtained and, finally, in section 7, some conclusions are drawn.

2. TIME AND FREQUENCY TOOLS FOR THE CHARACTERIZATION

In order to analyze the guitar sounds (*wav* files sampled at 44100 Hz) to obtain features to estimate the pitch and the string played, some initial processing stages are performed in the time and frequency domains. These are explained next.

2.1. Time domain analysis

The envelope of the notes of the guitar contains pieces of information about how the note under analysis has been played. We consider the envelope model proposed by Jensen [2]. To obtain the envelope, a Butterworth filter of order 5 and cutoff frequency 11 Hz is used. Before the filter, the audio samples are squared. After the filter, the signal is normalized so that the maximum is 1 and the samples under a 2.5% of the maximum amplitude are eliminated. Finally, the attack and release times are calculated. The attack time (T_a) lasts from the first sample that reaches a 10% of the maximum amplitude until the first sample that reaches a 85% of the maximum amplitude. The release time (T_r) is from the last sample that reaches a 70% of the maximum amplitude to the last one over the 10% of the maximum amplitude. The time between the attack time and the release time is called sustain time (T_s).

2.2. Frequency domain analysis

Two types of frequency analysis have been done, in order to obtain the different features that will be explained in section 4. In the first one, the time domain signal is analyzed using non-overlapping windows. This way, we will obtain, not only information about the frequency location and amplitudes of the fundamental harmonic and its partials, but also about their temporal evolution, as it will be explained in subsection 4.5. The localization of spectral peaks will be done using the method proposed in [3]. The size of the windows has been set to 16384 samples, this size is big enough to have a reasonable frequency resolution (2.7 Hz) and small enough to gather information about the temporal evolution (0.37 s).

The second frequency domain-analysis has been performed removing the samples that correspond to the attack time (T_a), (2.1) and using the following 65536 samples. During the attack time, there is a large amount of noise and spurious frequencies that seem not to provide any information about the played string, so these samples have been discarded. The number of samples has been chosen so that we have good frequency resolution (0.67 Hz) to obtain information about the inharmonicity [4], [5],with the restriction that all the sound files are long enough to have this number of samples.

3. PITCH ESTIMATION

To do the pitch estimation, the second frequency domain analysis has been considered. Once the peaks of the power spectrum have been found using the methods proposed in [3], they are sorted from minor to major frequency. Let $f_1 = f_{pitch}$ denote the fundamental frequency of the played note, the expected spectral peaks should be located around $f_1 = f_{pitch},\ f_2 = 2f_{pitch},\ f_3 = 3f_{pitch}$ and so on. Therefore, the minor frequency is considered as f_{pitch}, and it is checked if the subsequent spectral peaks follow, approximately, the expected harmonic relation nf_{pitch}. If this test is passed, this frequency is selected as the pitch. If, for any reason, the first harmonic is not present, then the expected harmonic relationship will not be fulfilled and, in order to identify the pitch, it is checked if the peak detected with with lowest frequency corresponds to the second harmonic: its frequency is taken as $2f_{pitch}$ and the expected harmonic relation is checked again.

In the analyzed sound files, the harmonic at the fundamental frequency has been always found, even for the lowest notes that can be played in a guitar. Figure 1 shows the spectrum of the lowest (E2) and the highest (E5) note for a classic guitar and figure 2 for an acoustic guitar. In all cases, it can be observed that the peak corresponding to the fundamental frequency of the played note is present, though they have different relative amplitude with respect the rest of the spectral peaks.

4. FEATURE EXTRACTION FOR THE CLASSIFICATION

In this section, the features that have been extracted in the time and frequency domains are presented.

4.1. Frequency centroid

Let $P(f_i)$ denote the discrete power spectrum of a signal, the frequency centroid, that can be interpreted as the center of mass of the power spectrum of the signal, is defined as

$$f_{cent} = \frac{\sum_i f_i P(f_i)}{\sum_i P(f_i)} \tag{1}$$

where i indexes the frequency bins of the discrete power spectrum and stands f_i for the frequency represented by the bin i. In order to calculate this frequency centroid, we have considered up to 3000 Hz because over that frequency the noise is dominant and there is almost no relevant frequency information regardless the pitch of the

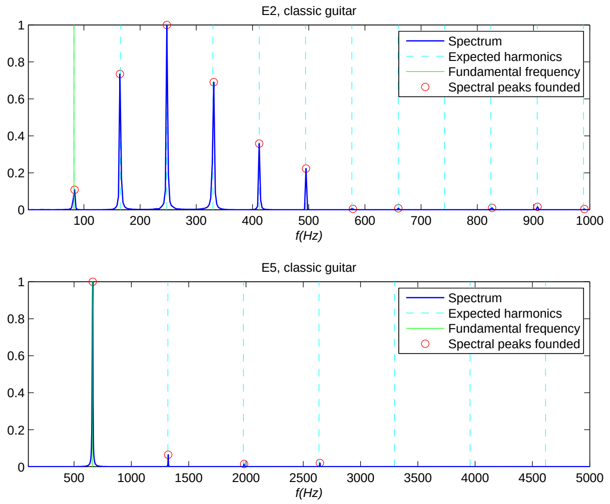

Fig. 1: Spectrum, expected harmonics, fundamental frequency and spectral peaks founded for the lowest (E2) and the highest (E5) notes of a classic guitar (extracted from RWC data base, '091CGAFM.wav').

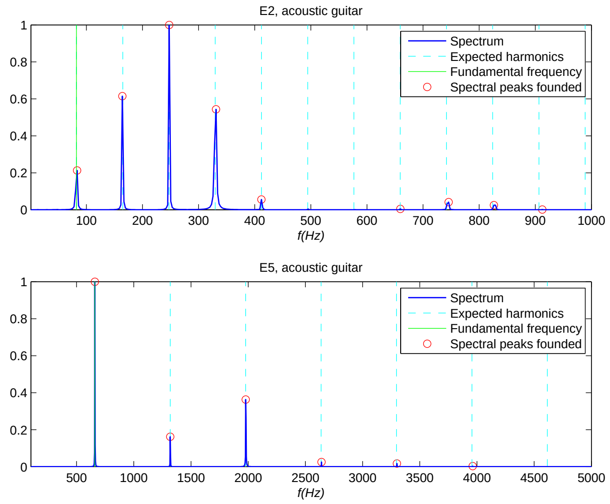

Fig. 2: Spectrum, expected harmonics, fundamental frequency and spectral peaks founded for the lowest (E2) and the highest (E5) notes of an acoustic guitar (extracted from RWC data base, '112AGAFM.wav').

note under analysis. Also, the f_1-adjusted centroid is considered [6] that is defined as the ratio of the centroid f_{cent} to the fundamental frequency f_1:

$$f_{cent,1} = \frac{f_{cent}}{f_1} \tag{2}$$

4.2. Nontonal spectral pattern

Another feature that has been extracted is the frequency centroid defined in equation (1) for the nontonal spectral pattern. The nontonal spectral pattern is the spectrum of the note where the partials produced by the harmonic vibration of the instrument strings have been removed [7]. The nonharmonic modes of vibration affect the color of the note and the sensation of the instrument that has generated that sound, therefore, it can be expected that it contains information about the played string.

4.3. Amplitude relationship between harmonics

The relationship between the relative amplitude of the fundamental frequency with respect to the partials is extracted. Specifically, the following ratios have been calculated $P(f_1)/P(f_2)$, $P(f_1)/P(f_3)$, $P(f_2)/P(f_3)$, $P(f_1)/P(f_4)$, where, in this case, f_i denotes the estimated frequency of the partial i and P denotes its estimated normalized amplitude using an exponential decay model, applied after the first window analysis. More specifically, the amplitude of the exponential decay model considered for the peaks of the partials detected in the frequency domain is divided by the maximum of the envelope of the amplitude (see section 2.1), then the following model is assumed for the evolution of the amplitude of each partial over the selected analysis windows:

$$P_{f_i}(t) = P(f_i)e^{-\frac{t}{\tau_i}} \tag{3}$$

where $P_{f_i}(t)$ describes the evolution of the amplitude of the partial i with the time and $P(f_i)$ represents the normalized amplitude of the exponential model. Taking the logarithm in eq. (3), we obtain a linear model in the parameters $\ln(P(f_i))$ and $\frac{1}{\tau_i}$ which are estimated using the least squares method.

4.4. Inharmonicity coefficient

An ideal string vibrates in a series of modes that are harmonics of a fundamental, but the real string, that has a limited length and a certain stiffness, is inharmonic. The inharmonicity of a string is the amount by which the actual mode frequencies differ from a harmonic series and it appears in the following relation:

$$f_n = nf_1\sqrt{1+n^2B} \tag{4}$$

where B is the inharmonicity coefficient. This coefficient depends on the physical characteristics of the string: length, radius and tension. In a guitar, the length of the plucked string will be different, depending on which is string has been used to play the note. Also, the strings are made of different materials and have different radius. Therefore, inharmonicity should contain information about the played string.

The difference between the harmonic frequencies and the inharmonic ones is very small, so the long spectrum of 65536 samples is used to have better frequency resolution. To estimate the inharmonicity coefficient B, the following steps are done: First, the fundamental frequency f_1 is estimated as the frequency around the theoretical-fundamental frequency (± 45 cents) where a maximum in the spectrum is found. Then, taking into account that the partials should be located slightly above the harmonic frequencies, to calculate f_n, the maximum in the frequency range $(nf_1, nf_1 + 30$ cents$)$ is searched and the frequency where the maximum is located is defined as f_n. Then, B is calculated using (4).

4.5. Exponential decay of the harmonics

The temporal evolution of the amplitude of the harmonics after the attack is assumed to be exponential. Using the windowed spectrum of the signal the amplitude and the time constant of the exponential is calculated for the fundamental harmonic and for the second and third partial, taking into account the considerations in section 4.3. To calculated them, the first window of 16384 samples has been discarded.

4.6. Relationships between attack, sustain and release times

The attack, sustain and release times of a sound contain information about the characteristics of that sound. Therefore, the following ratios have been also estimated: T_a/T_s, T_a/T_r and T_s/T_r, using the signal envelope and according to the descriptions given in section 2.1.

5. CLASSIFICATION

In this section, we show how we use the extracted features to identify the string plucked. The pitch of the note is known because it has been previously detected as explained in section 3. In our case, we deal with a large number of features, as shown, and we expect to use them in an efficient way in a multi-feature classifier.

A Fisher linear discriminant [8] has been implemented and a classification approach based on the Gaussian model [9] has been used. A simple strategy is implemented to classify between pairs of classes with the projected vectors using the Fisher criterion (one versus one classifier).

The Fisher method is settled to obtain a projection vector a, such that the projections $Y = a^T X$ generate two clouds of points, for the two classes of samples, such that their dissimilarity is as high as possible over the projection line. After the classifier is trained with a large number of known features vectors, the classification is performed on the base of a probabilistic approach.

The projections of the two classes are considered samples of Gaussian random variables [9]. Then, the class membership of an unknown sample can be determined using the probability that a projected sample lies further from the mode of the class distribution than the sample under test. To obtain this parameter, we evaluate the complementary error function and the result of the comparison of the probability measures of each pair of classes leads to the final classification given by the class that gives rise to the largest of the probability measures calculated. Such method is simple and computationally tractable for small number of classes, as our case.

6. RESULTS

The proposed algorithm has been tested using professional recorded samples of classic and acoustic guitars taken from the Musical Instrument Sound Data Base RWC-MDB-1-2001-W03 [10]. The set of guitar samples of each note in the database is composed as follows: all the notes are plucked at each possible string for three classic guitars (nylon strings) and three acoustic guitars (steel string), with three dynamics levels (*forte*, *mezzo*, *piano*) and different articulation methods. The articulation methods considered are Apoyando/Finger, Apoyando/Nail, Al aire/Finger and Al aire/Nail for the classic guitars and Apoyando/Finger, Apoyando/Pick, Al aire/Finger and Al aire/Pick for the acoustic guitars. This leads to a total of 36, 72 or 108 samples for each note, depending on whether a note can be played at one, two or three strings, respectively.

The results about the pitch detection are very good as expected in such a monophonic case. Actually, we obtain almost a 0% error rate. The discussion of the results is much more interesting for the estimation of the plucked string, this discussion follwos.

For the plucked string estimation it is assumed that the type of guitar (classic or acoustic) is known and also the pitch of the played note is known or estimated. After an analysis of the behavior of the extracted features regarding the string plucked to produce the sound of a certain note, we found that the behavior of the different features happens to be unstable or unpredictable in different notes. This behavior leads to use different features for the detection of the plucked string for different notes and, also, to obtain specific projection vectors and decision parameters for the analysis of each target note.

After a long revision of the behavior of the classifier, carried on by manually selecting a varying number of descriptors among the available ones, we decided to use only three at a time, because a systematic improvement of the system performance by using a larger amount of features was not found, rather, the features are selected among the ones that seem to have certain discrimination capabilities.

In the following figures, at the top, the symbolized scatter plot of each descriptor is represented, at the bottom, its categorized box-whiskers plot is shown, where the (blue) box and the (red) line represent the first and the third quartile and the median value, respectively. The (black) whiskers represent 1.5 times the interquartile range (no-outlier range) from the limits of the box (within minimum or maximum of the distribution) and the (red) crosses represent the remaining samples not included in the whiskers range (outliers).

The same descriptor can show very different behavior for different notes. For example, observe the behavior of the nontonal spectrum centroid of Figs. 3 and 4. In the case of the note $G3$ played on acoustic guitars (Fig. 3), the distributions of the descriptor for the different strings are characterized by a high degree of overlapping, with similar median values. Conversely, in the case of the note $G^{\#}3$ played on acoustic guitars (Fig. 4), the distributions are scarcely overlapped, with well distinct medians and distributions. Note that $G3$ and $G^{\#}3$ are only half tone far away and involve the same three strings.

Also, two different descriptors can behave oppositely for the same note. For the note $G^{\#}3$, played on acoustic guitars (Fig. 5), the amplitude of the exponential of the first harmonic reveals a very poor discriminating capability, showing nearly overlapping distributions, this result contrasts with the one provided for the nontonal spectrum centroid descriptor computed on the same note, shown in Fig. 4.

Other descriptors works fine, as the f_1-adjusted centroid, computed for the note $D4$, played on acoustic guitars (Fig. 6) or the ratio between the attack and sustain times computed for the note $C^{\#}4$, played on classic guitars (Fig. 7), both showing a nearly null overlapping in the distributions of the different strings, with different median and with low overlap of their dispersions. This fact reveals a high importance of such descriptors for the overall discrimination capability of the classifier.

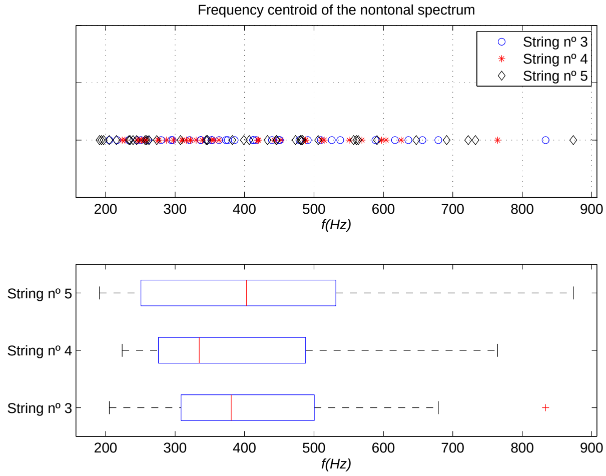

Fig. 3: Graph of the behavior of the nontonal spectrum centroid for the note $G3$ played on acoustic guitars.

From a intensive analysis of the results of each descriptor on a wide range of note samples, it emerges that the behavior of the features has a high degree of dependence on the note and can optimally work only in specific conditions.

Note that different notes can involve the use of different strings made with different material. Such aspect is particularly evident in the classic guitar, however, this fact has not been found to be useful for the selection of the features for the identification of the plucked string. In Figs. 8 and 9, the ratios between the attack and the sustain time and between the sustain and the release time, respectively computed for the notes $E4$ and $D3$, are shown. Such notes are played on classic guitars, where the two groups of three strings are made by the same material (in bronze for the note $D3$ and in nylon for the note $E4$). It is clear that these descriptors are not able to perform the desired discrimination in the two notes due to the the high degree of similarity of the envelope parameters (attack, sustain and release times) mainly due to the use of the

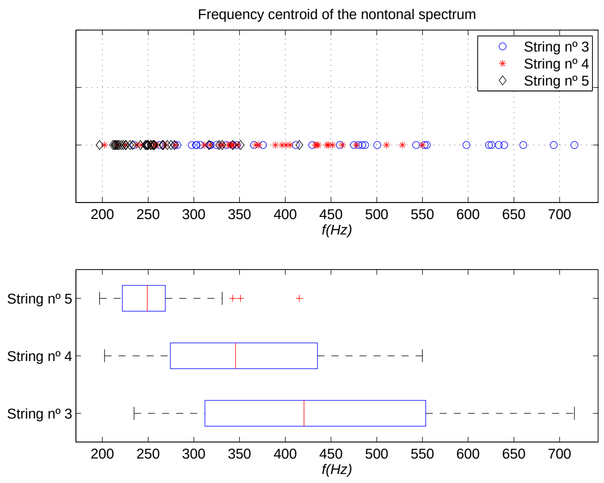

Fig. 4: Graph of the behavior of the nontonal spectrum centroid for the note $G^{\#}3$ played on acoustic guitars.

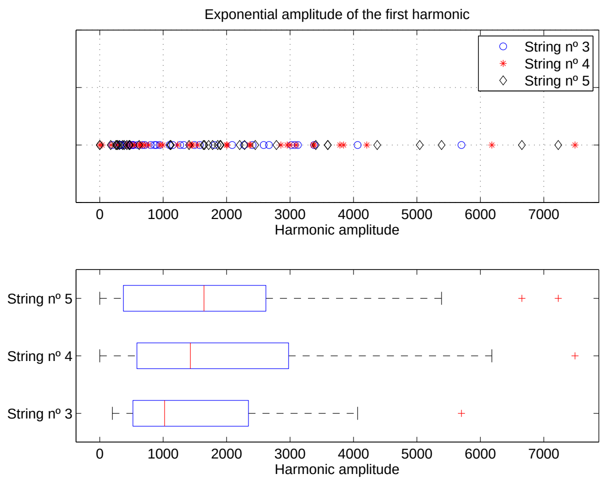

Fig. 5: Graph of the behavior of the amplitude of the exponential of the first harmonic for the note $G^{\#}3$, played on acoustic guitars.

same material.

Conversely, for the note B3, involving the strings 2nd and 3rd, in nylon, and the 4th, in bronze, the different responses of the material imply high discrimination capabilities of the 4th string with respect to the 2nd and the 3rd strings (Fig. 10)

Hence, it is clear that the descriptors heavily depend on the tone and there has not been found a systematic and predictable behavior of the descriptor for the given data. This leads to a high heterogeneity of the results of the classifier, based on the same features.

In Table 1, some results for both the acoustic and classic guitars are shown. Some example notes are presented, each of them involving two or three strings, from the thinest to the thickest one. Three descriptors have been employed in the classifier, among the available ones: the frequency centroid, the amplitude of the exponential of the third harmonic and the ratio between the sustain and the release times. Such choice is justified by a large search among all the possible combinations of descriptors in groups of three, and the analysis of the results.

The same subset of descriptors can perform suitable and return an error rate of 7% approximatively, as in the case of the note $B3$ played on the classic guitar, or even an error rate of 0% (on the available dataset) as in the case of the 2-string note $B2$ played on the acoustic guitar, and a value of 73%, approximately, for the note $G3$, played on classic guitars. Moreover, the same note, $G3$, that involves the strings 3rd, 4th and 5th, made of different materials in classic guitars and of the same phosphorous bronze material in acoustic ones, shows two extremely different results when played in classic guitars ($\sim$ 73%) and in acoustic ones ($\sim$ 13%). The results are referred to the classifiers described trained with 93 elements (31 for each string) and the test is performed with the remaining 15 samples (5 for each string).

A deeper analysis can be done on the results obtained, if the specific guitar model for a certain type of guitar is taken into account. The fact that each sample is different from the others, regarding the different guitar models, playing styles and dynamics, leads to high sensibility of the system with the training data set. That is, the classifier is trained with a number of samples that are related to some playing conditions, these conditions are substantially different among them, even if the same note is played and the same strings are plucked. These different conditions happen between the different training samples and between the training samples and the test samples. A number of samples of the same note played under the same conditions would be necessary to improve the performance of the system and reduce the mentioned sensibility.

In Table 2, the classification error rates of the note $D3$, played on the classic guitar, are shown. The error rates in

Note	Guitar Type	Classification Error rate			
		n	n+1	n+2	Global
B2	**Classic**	0.0 %	80.0 %	-	**40.0** %
	Acoustic	0.0 %	0.0 %	-	**0.0** %
D3	**Classic**	0.0 %	40.0 %	0.0 %	**13.3** %
	Acoustic	40.0 %	20.0 %	0.0 %	**20.0** %
G3	**Classic**	40.0 %	100.0 %	80.0 %	**73.3** %
	Acoustic	0.0 %	20.0 %	20.0 %	**13.3** %
B3	**Classic**	20.0 %	0.0 %	0.0 %	**6.7** %
	Acoustic	20.0 %	20.0 %	0.0 %	**13.3** %
D4	**Classic**	20.0 %	0.0 %	0.0 %	**6.7** %
	Acoustic	40.0 %	20.0 %	60.0 %	**40.0** %
A4	**Classic**	40.0 %	20.0 %	-	**30.0** %
	Acoustic	20.0 %	20.0 %	-	**20.0** %

Table 1: Error rates for 2- and 3-strings notes $B2$, $D3$, $G3$, $B3$, $D4$ and $A4$, played both in classic and acoustic guitars. Error rates are shown for each of the strings involved, where n is the finest string. Such results have been obtained by employing the following three descriptors: frequency centroid, amplitude of the exponential of the third harmonic and the ratio T_s/T_r. Global error rates are shown too.

the first row ($D3$) are referred to the complete set of three guitar models employed, while the ones in the second row ($D3$*) are obtained by removing one guitar model. In this case, there are 24 samples of each string, 20 samples are used in the training data set and 4 for the test.

7. CONCLUSIONS

In this paper, we have presented a method to detect not only the pitch but also to identify the plucked string in classic and acoustic guitars. The proposed method is based on the time-domain and the frequency-domain analysis of the sound.

The pitch estimation method is based on the analysis of the spectral peaks of the signal and it is almost error free in our monophonic context.

The identification of the plucked string has been found to be a much more difficult problem. Several features of the sound has been extracted and their discrimination capabilities have been observed: frequency centroid, non-tonal spectral pattern frequency centroid, amplitude relationship between harmonics, etc. No one of the extracted features provides a simple criterium to decide which string has been plucked to produce a certain note, even when different strings of different materials are considered. The observation of the behavior of the different descriptors reveals that the information provided for each descriptor depends on the guitar type and on the specific note.

In order to use several features at a time to improve the performance of the identification of the plucked string, a Fisher linear discriminant has been implemented. Again, the classification error rate obtained depends on the guitar type and on the note that has been played, but also, as expected, on the set of features considered by the discriminant to make the decisions. Therefore, further research should be done in order to decide the optimum set of descriptors to be used for each note to detect the plucked string and other features should be extracted.

Also, a larger dataset for training and testing would be necessary to build more robust classifiers and to obtain more robust parameters of the models of the descriptors. Specifically, a number of samples of each note played under each of the possible conditions would be required, although this would multiply the size of the resulting database and the analysis of the data would be more cumbersome.

Note	**Classification Error rate**			
	4th	5th	6th	Global
D3	0.0 %	40.0 %	0.0 %	**13.3** %
D3*	0.0 %	0.0 %	0.0 %	**0.0** %

Table 2: Error rates for note $D3$, played in classic guitars. For $D3$, the classifier is applied to the whole set of guitar models, while for $D3$*, one of the guitar model has been removed from the dataset. Such results have been obtained by employing the same descriptors used before: frequency centroid, amplitude of the exponential of the third harmonic and the ratio T_s/T_r. Global error rates are shown too.

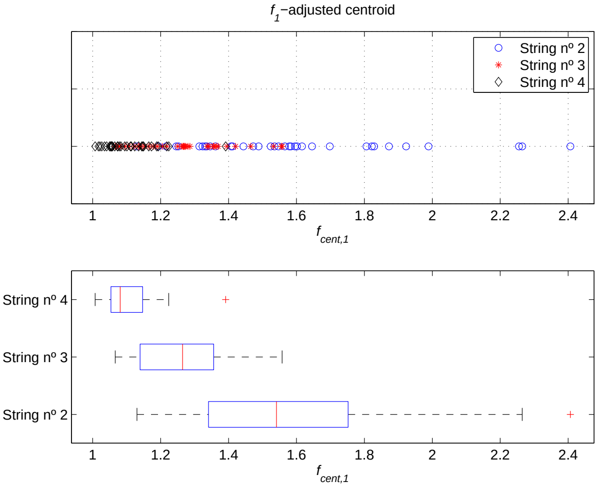

Fig. 6: Graph of the behavior of the f_1-adjusted centroid, for the note $D4$, played on acoustic guitars.

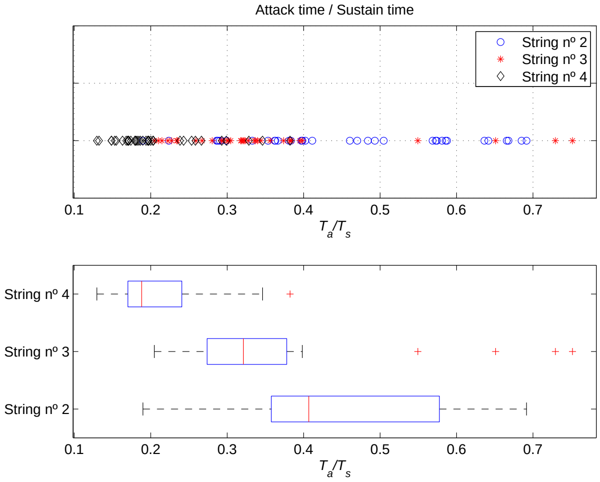

Fig. 7: Graph of the behavior of the ratio between the attack and the sustain times, for the note $C^{\#}4$, played on classic guitars.

8. ACKNOWLEDGMENTS

This work has been funded by the Ministerio de Educación y Ciencia of the Spanish Government under Project No. TSI2007-61181.

9. REFERENCES

[1] Traube, C., Smith, J. O., 'Estimating the plucking point on a guitar string', Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), December 2000.

[2] Jensen, K., 'Envelope Model of Isolated Musical Sounds', Proceedings of DAFx99, 1999.

[3] Tardón, L. J. Barbancho, I. 'Spectral Peaks Detection for Music and Audio Applications', IEEE Transactions on Audio, Speech and Language Processing, under consideration for publication 2008.

[4] Godsill, S., Davy, M, 'Bayesian computational models for inharmonicity in musical instruments', IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 2005.

[5] Karjalainen M., Järveläinen, H., 'Is inharmonicity perceivable in the acoustic guitar?', ForumAcusticum, 2005.

[6] Schubert, E., Wolfe, J. 'Does timbral brightness scale with frequency and spectral centroid?', Acta Acustica United with Acustica, vol. 92, pp.820-825, 2006.

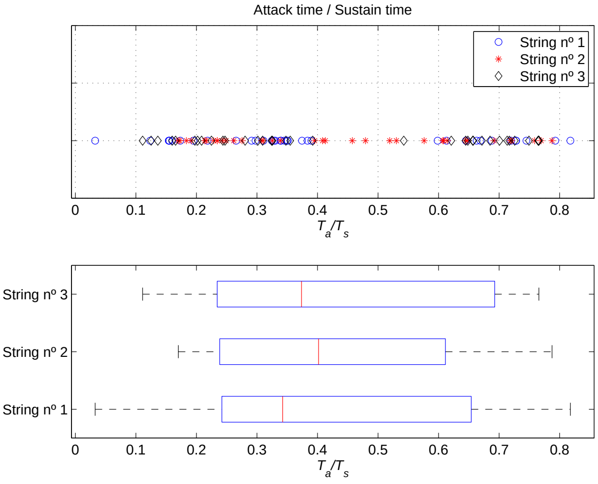

Fig. 8: Graph of the behavior of the ratio between the attack and the sustain times for the note $E4$, played on classic guitars.

[7] Fragoulis, D., Papaodysseus, C., Exarhos, M., Roussopoulos, G., 'Automated classification of piano-guitar notes', IEEE Transactions on Audio, Speech and Language Processing, vol. 14, no. 3, pp. 1040-1050, May 2006.

[8] Duda,R.O., Hart, P.E., Stork,D.G. , 'Pattern Classification (2nd Edition)', *Wiley-Interscience*, October 2000.

[9] Stark, H., Woods, J. W., 'Probability and Random Processes with Applications to Signal Processing (3rd Edition)', *Prentice Hall*, August 2001.

[10] Goto, M., 'Development of the RWC Music Database', Proceedings of the 18th International Congress on Acoustics, April 2004.

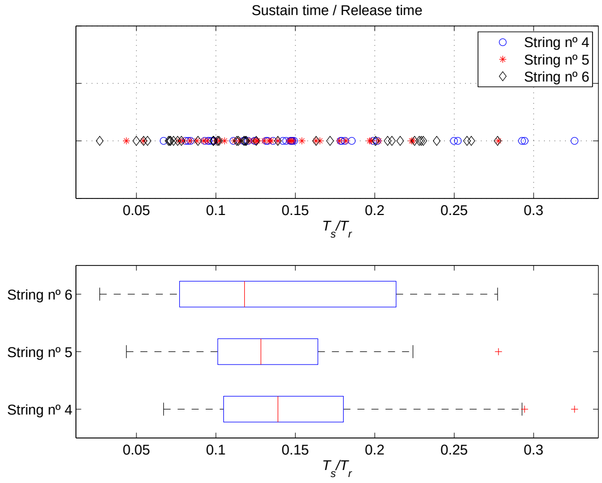

Fig. 9: Graph of the behavior of the ratio between the sustain and the release times for the note $D3$, played on classic guitars.

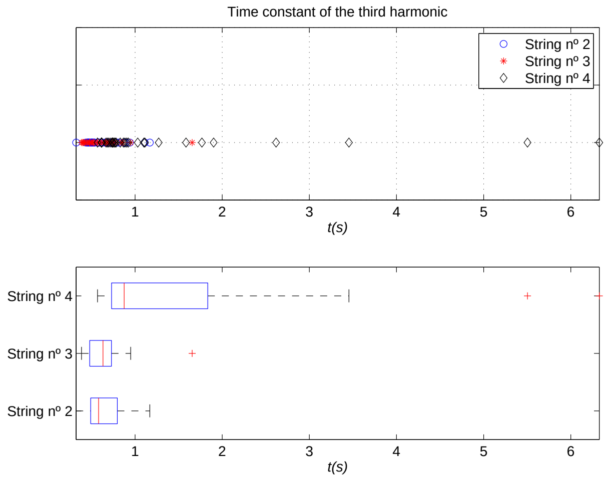

Fig. 10: Graph of the behavior of the time constant for the note $B3$, played on classic guitars.